GALIP: Generative Adversarial CLIPs for Text-to-Image Synthesis

Year: 2023 | Citations: 133 | Authors: Ming Tao, Bingkun Bao, Hao Tang, Changsheng Xu

---

Abstract

Synthesizing high-fidelity complex images from text is challenging. Based on large pretraining, the autoregressive and diffusion models can synthesize photo-realistic images. Although these large models have shown notable progress, there remain three flaws. 1) These models require tremendous training data and parameters to achieve good performance. 2) The multi-step generation design slows the image synthesis process heavily. 3) The synthesized visual features are challenging to control and require delicately designed prompts. To enable high-quality, efficient, fast, and controllable text-to-image synthesis, we propose Generative Adversarial CLIPs, namely GALIP. GALIP leverages the powerful pretrained CLIP model both in the discriminator and generator. Specifically, we propose a CLIP-based discriminator. The complex scene understanding ability of CLIP enables the discriminator to accurately assess the image quality. Furthermore, we propose a CLIP-empowered generator that induces the visual concepts from CLIP through bridge features and prompts. The CLIP-integrated generator and discriminator boost training efficiency, and as a result, our model only requires about 3% training data and 6% learnable parameters, achieving comparable results to large pretrained autoregressive and diffusion models. Moreover, our model achieves ~120×faster synthesis speed and inherits the smooth latent space from GAN. The extensive experimental results demonstrate the excellent performance of our GALIP. Code is available at https://github.com/tobran/GALIP.